PCT WELTORGANISATION FÜR GEISTIGES EIGENTUM
Internationales Büro
INTERNATIONALE ANMELDUNG VERÖFFENTLICHT NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT)

(51) Internationale Patentklassifikation 6 : G10L 5/06 — A1

(11) Internationale Veröffentlichungsnummer: **WO 98/11534**

(43) Internationales Veröffentlichungsdatum: 19. März 1998 (19.03.98)

(21) Internationales Aktenzeichen: PCT/DE97/02016

(22) Internationales Anmeldedatum: 10. September 1997 (10.09.97)

(30) Prioritätsdaten:
196 36 739.5 10. September 1996 (10.09.96) DE
196 40 586.6 1. Oktober 1996 (01.10.96) DE

(71) Anmelder *(für alle Bestimmungsstaaten ausser US)*: SIEMENS AKTIENGESELLSCHAFT [DE/DE]; Wittelsbacherplatz 2, D-80333 München (DE).

(72) Erfinder; und
(75) Erfinder/Anmelder *(nur für US)*: BUB, Udo [DE/DE]; Klarweinstrasse 18, D-81247 München (DE). HÖGE, Harald [DE/DE]; Obertaxetweg 6 B, D-82131 Gauting (DE). KÖHLER, Joachim [DE/DE]; Gudrunstrasse 14, D-80634 München (DE).

(81) Bestimmungsstaaten: BR, CN, JP, US, europäisches Patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

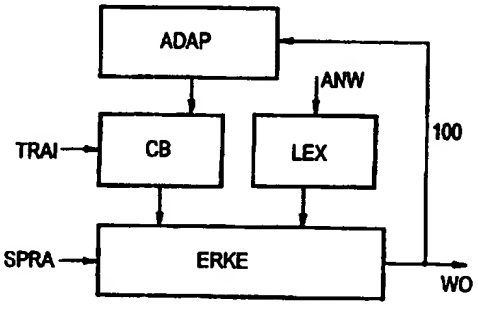
Veröffentlicht
Mit internationalem Recherchenbericht.
Vor Ablauf der für Änderungen der Ansprüche zugelassenen Frist. Veröffentlichung wird wiederholt falls Änderungen eintreffen.

(54) Title: PROCESS FOR ADAPTATION OF A HIDDEN MARKOV SOUND MODEL IN A SPEECH RECOGNITION SYSTEM

(54) Bezeichnung: VERFAHREN ZUR ANPASSUNG EINES HIDDEN-MARKOV-LAUTMODELLES IN EINEM SPRACHERKENNUNGSSYSTEM

(57) Abstract

Process for adaptation of a hidden Markov sound model in a speech recognition system. This invention concerns a process for adapting a generally available code book (CB) for special applications with a speech recognition system of the hidden Markov sound model. These applications are defined by a lexicon (LEX) changed by the user. The adaption (ADAP) is done during operation and occurs by means of a displacement of the stored midpoint vector of the probability density distributions of hidden Markov models, in the direction of a known feature vector of sound expressions and in relationship to the hidden Markov models specially used. In comparison to current practices, the invention has the advantage that it is done on-line and that it has a very high recognition rate with little computational expenditure. In addition, the training expenditure for special sound models for corresponding applications is avoided. By using special hidden Markov models from multilingual phonemes, in which sound similarities across various languages are used, automatic adaptation to foreign languages can follow. Both language-specific and language-dependent characteristics are taken into account in this method for acoustic phonetic modelling to determine the probability densities for different hidden Markov sound models in different languages.



(57) Zusammenfassung

Mit der Erfindung wird ein allgemein mit einem Spracherkennungssystem zur Verfügung gestelltes Codebuch (CB) von hidden-Markov-Lautmodellen für spezielle Anwendungsfälle adaptiert. Diese Anwendungsfälle werden dabei durch ein vom Anwender verändertes Lexikon (LEX) der Applikation definiert. Die Adaption (ADAP) erfolgt während des Betriebs und geschieht durch eine Verschiebung des gespeicherten Mittelpunktsvektors der Wahrscheinlichkeitsdichteverteilungen von hidden-Markov-Modellen, in Richtung eines erkannten Merkmalsvektors von Lautäußerungen und in Bezug auf die speziell verwendeten hidden-Markov-Modelle. Gegenüber gängigen Verfahren hat die Erfindung den Vorteil, daß sie On-Line erfolgt und daß sie eine sehr hohe Erkennungsrate bei einem geringen Rechenaufwand gewährleistet. Weiterhin wird der Aufwand für das Training von speziellen Lautmodellen für entsprechende Einsatzfälle vermieden. Durch Anwendung spezieller hidden-Markov-Modelle aus multilingualen Phonemen, bei denen die Ähnlichkeiten von Lauten über verschiedene Sprachen hinweg ausgenutzt wird, kann eine automatische Adaption an Fremdsprachen erfolgen. Bei der dabei verwendeten Methode zur akustisch phonetischen Modellierung werden sowohl sprachspezifische als auch sprachunabhängige Eigenschaften bei der Zusammenfassung der Wahrscheinlichkeitsdichten für unterschiedliche hidden-Markov-Lautmodelle in verschiedenen Sprachen berücksichtigt.

LEDIGLICH ZUR INFORMATION

Codes zur Identifizierung von PCT-Vertragsstaaten auf den Kopfbögen der Schriften, die internationale Anmeldungen gemäss dem PCT veröffentlichen.

AL	Albanien	ES	Spanien	LS	Lesotho	SI	Slowenien
AM	Armenien	FI	Finnland	LT	Litauen	SK	Slowakei
AT	Österreich	FR	Frankreich	LU	Luxemburg	SN	Senegal
AU	Australien	GA	Gabun	LV	Lettland	SZ	Swasiland
AZ	Aserbaidschan	GB	Vereinigtes Königreich	MC	Monaco	TD	Tschad
BA	Bosnien-Herzegowina	GE	Georgien	MD	Republik Moldau	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagaskar	TJ	Tadschikistan
BE	Belgien	GN	Guinea	MK	Die ehemalige jugoslawische Republik Mazedonien	TM	Turkmenistan
BF	Burkina Faso	GR	Griechenland			TR	Türkei
BG	Bulgarien	HU	Ungarn	ML	Mali	TT	Trinidad und Tobago
BJ	Benin	IE	Irland	MN	Mongolei	UA	Ukraine
BR	Brasilien	IL	Israel	MR	Mauretanien	UG	Uganda
BY	Belarus	IS	Island	MW	Malawi	US	Vereinigte Staaten von Amerika
CA	Kanada	IT	Italien	MX	Mexiko		
CF	Zentralafrikanische Republik	JP	Japan	NE	Niger	UZ	Usbekistan
CG	Kongo	KE	Kenia	NL	Niederlande	VN	Vietnam
CH	Schweiz	KG	Kirgisistan	NO	Norwegen	YU	Jugoslawien
CI	Côte d'Ivoire	KP	Demokratische Volksrepublik Korea	NZ	Neuseeland	ZW	Zimbabwe
CM	Kamerun			PL	Polen		
CN	China	KR	Republik Korea	PT	Portugal		
CU	Kuba	KZ	Kasachstan	RO	Rumänien		
CZ	Tschechische Republik	LC	St. Lucia	RU	Russische Föderation		
DE	Deutschland	LI	Liechtenstein	SD	Sudan		
DK	Dänemark	LK	Sri Lanka	SE	Schweden		
EE	Estland	LR	Liberia	SG	Singapur		

Beschreibung

Verfahren zur Anpassung eines hidden-Markov-Lautmodelles in einem Spracherkennungssystem

Die Erfindung bezieht sich auf ein Verfahren zur Anpassung von hidden-Markov-Lautmodellen an Betriebserfordernisse eines Spracherkennungssystems, insbesondere unter Verwendung speziell gebildeter Mehrsprachen hidden-Markov-Lautmodelle, die an eine Einsatzsprache angepaßt werden.

Ein Spracherkennungssystem greift im wesentlichen auf zwei unabhängige Wissensquellen zu. Zum einen ist dies ein phonetisches Lexikon, mit dem das zu erkennende Vokabular als Wortschatz festgelegt wird. Dort werden beispielsweise die ASCII-Strings der einzelnen zu erkennenden Worte, sowie deren phonetische Umschrift gespeichert. Ebenso wird durch dieses Lexikon eine sogenannte "Task" vorgegeben.
Zum anderen ist dies ein Codebuch, das die Parameter der Hidden-Markov-Lautmodelle (HMM) und damit insbesondere die Mittelpunkte der zu Erkennungssegmenten gehörigen Wahrscheinlichkeitsdichteverteilungen enthält.

Die beste Leistung eines Spracherkennungssystems ist zu beobachten, wenn das HMM-Codebuch optimal auf das Lexikon abgestimmt ist. Dies ist der Fall, wenn das HMM-Codebuch zusammen mit demjenigen Lexikon betrieben wird, mit dem dieses HMM-Codebuch auch eingangs durch Training erstellt wurde. Kann dies nicht gewährleistet werden, so ist ein Leistungseinbruch feststellbar.

Bei Spracherkennungssystemen, wie sie beispielsweise in Vermittlungssystemen eingesetzt werden, tritt häufig das Problem auf, daß der eingangs trainierte Wortschatz, mit welchem dieses System ausgeliefert wird im Betrieb durch den Kunden abgeändert wird. Dies hat in der Regel zur Folge, daß nun bei den neuen Wörtern des Lexikons Koartikulationen zwischen Pho-

nemen auftreten, die vorher nicht trainiert werden konnten. Es besteht nun also ein "Mismatch" zwischen Lexikon und HMM-Codebuch, was zu einer verschlechterten Erkennungsleistung im praktischen Betrieb führt.

Ein praktisches Beispiel für eine solche Situation wäre ein telefonisches Vermittlungssystem einer Firma, das die Namen der Mitarbeiter versteht und den Verbindungswunsch eines Anrufers durch dessen Spracheingabe automatisch erkennt und den Anruf an die entsprechende Nebenstelle weiterleitet (Call-by-Name). Im Lexikon sind also die Namen der Mitarbeiter gespeichert. Durch Fluktuation werden sich die Namen immer wieder ändern und das System wird somit aus den genannten Gründen eine unbefriedigende Erkennungsleistung vorweisen.

Um eine möglichst hohe Erkennungsleistung eines Spracherkennungssystems unter den geschilderten Einsatzbedingungen zu gewährleisten, ist es also erforderlich, eine Anpassung (Adaption) des zugrundeliegenden HMM-Codebuchs dieses Erkennungssystems an die neu gegebene Task, die durch das veränderte Lexikon gegeben wurde, durchzuführen. Aus dem Stand der Technik sind unterschiedliche Verfahren zur Lösung dieses Problems bekannt. Aus [1] ist eine Lösung bekannt, bei der vorgeschlagen wird, ein Nachtraining zur Anpassung des Codebuchs an das Lexikon durchzuführen. Diese Vorgehensweise hat den Nachteil, daß im allgemeinen das Vokabular der Endanwendung zum Trainingszeitpunkt nur teilweise bekannt ist. Falls nun zu einem späteren Zeitpunkt das Nachtraining gestartet werden muß, so müssen alle potentiell benötigten akustischen Modelle eines neuen Vokabulars bereitgehalten werden, was unwirtschaftlich ist und praktisch schwer durchführbar wäre.

Aus [2] ist ein sogenannter MAP-Algorithmus (Maximum a Posteriori) zum Adaptieren der akustischen Modelle durch den Anwender auf Basis eines bestimmten Satzes von Sprachproben bekannt. Hierbei muß der Erwerber des Spracherkennungssystems Sprachproben von mehreren Sprechern zur Verfügung stellen.

Die Umadaption des Codebuches erfolgt dabei durch überwachtes Lernen, d. h. daß dem System die korrekte Transliteration einer Äußerung mitgeteilt werden muß. Die hierbei erforderlichen komplizierten Arbeitsschritte sind einem Kunden nicht zuzumuten.

Beide Lösungen aus dem Stand der Technik haben den gemeinsamen Nachteil, daß sie lediglich Off-Line ablaufen. Für eine HMM Codebuchadaption muß also das laufende System außer Betrieb genommen werden, damit die neuen Parameter, d. h. die entsprechenden Erkennungseinheiten in das System eingespielt werden können. Weiterhin erfordern die Vorgänge des Trainings und des Adaptierens eine große Zeit für die Einarbeitung und Durchführung, was einen finanziellen Nachteil für Erwerber des Systems bedeutet. Häufig wird deshalb bei Auslieferung des Produkts ein Ausgangscodebuch für die HMM bereitgestellt. Aus dem Stand der Technik bieten sich zwei Trainingsstrategien hierfür an.

Einerseits kann das Codebuch auf Basis eines phonetisch ausgeglichenen Trainingsdatensatzes generiert werden. Derartige Codebücher bieten den Vorteil, daß sie mit allen denkbaren Anwendungsfällen von unbekannten Aufgaben ("Tasks") fertig werden, da sie keine Erkennungseinheiten bevorzugen. Andererseits kann wenn möglich ein Spezialistencodebuch trainiert werden. Dabei wird das Spracherkennungssystem exakt auf denselben Wortschatz trainiert, welcher in der Endapplikation eine Rolle spielt. Hierdurch wird eine höhere Erkennungsrate für die Spezialanwendung hauptsächlich dadurch erzielt, daß das Spracherkennungssystem von Koartikulationen Gebrauch machen kann, welche es schon in der Trainingsphase trainiert bekam. Für Anwendungen bei denen sich das Lexikon ändert, zeigen solche Spezialistencodebücher aber schlechtere Leistungen.

Ist das Lexikon und damit der Wortschatz der Endanwendung, wie in dem für die Erfindung relevanten Fall, veränderbar,

oder zum Trainingszeitpunkt gar gänzlich unbekannt, so sind Hersteller folglich häufig bestrebt, ein möglichst allgemein gehaltenes Codebuch in ihre Spracherkennungssysteme einzuarbeiten.

Weiterhin besteht ein großes Problem darin, daß für jede Sprache in welcher die Spracherkennungstechnologie eingeführt werden soll, neue akustisch phonetische Modelle trainiert werden müssen, um eine Länderanpassung durchführen zu können. Meistens werden bei Spracherkennungssystemen HMM zur Modellierung der sprachspezifischen Laute verwendet. Aus diesen statistisch modellierten Lautmodellen werden im Anschluß akustische Wortmodelle zusammengefügt, welche während eines Suchprozesses beim Spracherkennungsvorgang erkannt werden. Zum Training dieser Lautmodelle werden sehr umfangreiche Sprachdatenbanken benötigt, deren Sammlung und Aufbereitung einen äußerst kosten- und zeitintensiven Prozeß darstellt. Hierdurch entstehen Nachteile bei der Portierung einer Spracherkennungstechnologie von einer Sprache in eine weitere Sprache, da die Erstellung einer neuen Sprachdatenbank einerseits eine Verteuerung des Produktes bedeutet und andererseits eine zeitliche Verzögerung bei der Markteinführung bedingt.

In gängigen erwerbbaren Spracherkennungssystemen werden ausschließlich sprachspezifische Modelle verwendet. Zur Portierung dieser Systeme in eine neue Sprache werden umfangreiche Sprachdatenbanken gesammelt und aufbereitet. Anschließend werden die Lautmodelle für die neue Sprache mit diesen gesammelten Sprachdaten von Grund auf neu trainiert.

Um den Aufwand und die Zeitverzögerung bei der Portierung von Spracherkennungssystemen in unterschiedliche Sprachen zu verringern, sollte also untersucht werden, ob einzelne Lautmodelle für die Verwendung in verschiedenen Sprachen geeignet sind. Hierzu gibt es in [4] bereits Ansätze mehrsprachige Lautmodelle zu erstellen und diese bei der Spracherkennung in

den jeweiligen Sprachen einzusetzen. Dort werden auch die Begriffe Poly- und Monophoneme eingeführt. Wobei Polyphoneme Laute bedeuten, deren Lautbildungseigenschaften über mehrere Sprachen hinweg ähnlich genug sind, um gleichgesetzt zu werden. Mit Monophonemen werden Laute bezeichnet, welche sprachspezifische Eigenschaften aufweisen. Um für solche Entwicklungsarbeiten und Untersuchungen nicht jedesmal neue Sprachdatenbanken trainieren zu müssen, stehen solche schon als Standard zur Verfügung [8], [6], [9]. Aus [10] ist es bekannt vorhandene mehrsprachige Modelle zum Segmentieren der Sprachdaten in einer Zielsprache zu verwenden. Das Training der Lautmodellle wird dann in der Zielspache durchgeführt. Ein weiterer Stand der Technik zur mehrsprachigen Verwendung von Lautmodellen ist nicht bekannt.

Die der Erfindung zugrundeliegende Aufgabe besteht also darin, ein Verfahren zur Anpassung eines HMM in einem Spracherkennungssystem anzugeben, bei dem die Anpassung während des laufenden Betriebs des Spracherkennungssystems erfolgt. Insbesondere sollen durch die Anpassung die oben beschriebenen Komplikationen kompensiert werden, welche sich aus der Änderung des Lexikons und damit der Task ergeben.

Diese Aufgabe wird gemäß den Merkmalen des Patentanspruches 1 gelöst.

Eine weitere Aufgabe besteht demnach darin, ein Verfahren zur Bildung und Adaption spezieller mehrsprachenverwendbarer HMM in einem Spracherkennungssysten anzugeben, durch welches der Portierungsaufwand von Spracherkennungssystemen in eine andere Sprache minimiert wird, indem die Parameter in einem multilingualen Spracherkennungssytem reduziert werden.

Diese Aufgabe wird gemäß den Merkmalen des Patentanspruches 8 gelöst.

Weiterbildungen der Erfindung ergeben sich aus den abhängigen Ansprüchen.

Der erfindungsgemäße Weg sieht es dazu vor, ein allgemein gehaltenes Codebuch, welches beispielsweise HMM enthält, die für mehrere Sprachen gemeinsam Verwendung finden, als Saatmodell zu verwenden und es im laufenden Betrieb bei einem veränderten Lexikon an dieses neue Lexikon anzupassen.

Besonders vorteilhaft wird durch das Verfahren eine Anpassung im Betrieb dadurch erreicht, daß ein bereits erkannter Merkmalsvektor einer Lautäußerung zu einer Verschiebung des gespeicherten Mittelpunktsvektors im HMM-Codebuch führt, indem mittels eines Anpassungsfaktors im Betrieb nach dem Erkennen des Wortes oder der Lautfolge, eine Verschiebung des Mittelpunktes der Wahrscheinlichkeitsverteilung der hidden-Markov-Modelle in Richtung des erkannten Merkmalsvektors erfolgt. Die Lernrate kann dabei durch den Anpassungsfaktor beliebig eingestellt werden.

Vorteilhaft kann beim Verfahren die Zuordnung der Merkmalsvektoren zu den HMM mit Standardverfahren, wie dem Viterbi-Algorithmus durchgeführt werden. Durch Anwendung des Viterbi-Algorithmus liegt nach Erkennung eine eindeutige Zuordnung der Merkmalsvektoren zu den gespeicherten Mittelpunktsvektoren des HMM Codebuchs vor.

Besonders vorteilhaft werden die anzupassenden und zu erkennenden Lautmodelle in einem standardisierten HMM-Codebuch zur Verfügung gehalten, welches als Grundlage für alle anzupassenden Praxismodelle dienen kann und somit für alle anzupassenden Systeme nur einmal bei der Erstellung trainiert, bzw. in Form eines Codebuches mit Mehrsprachen-HMM bereitgestellt werden muß.

Besonders vorteilhaft erfolgt die Anpassung des Schwerpunktvektors an den erkannten Merkmalsvektor bei Laplace- und

Gauß- Wahrscheinlichkeitsdichteverteilungen der hidden-Markov-Modelle mit den speziell angegebenen Gleichungen, da damit ein vergleichsweise geringer Rechenaufwand verbunden ist.

Vorteilhaft wird beim aufgezeigten Verfahren eine noch höhere Erkennungsrate erzielt, wenn im Fall einer unsicher erkannten Lautäußerung diese komplett zurückgewiesen wird und keine Anpassung erfolgt.

Besonders vorteilhaft wird bei der Zurückweisung die Anzahl der Lauthypotesen nach der Viterbi-Suche und deren zugehörige Trefferraten der jeweiligen Hypothesen in bezug auf die Äußerung berücksichtigt. Die Zurückweisung wird in diesem Fall von den Unterschieden zwischen den Trefferraten abhängig gemacht, da diese Unterschiede eine Qualitätsangabe für die Güte der gefundenen Lösung darstellen. Bevorzugt kann bei großen Unterschieden keine Zurückweisung erfolgen und bei kleinen Unterschieden muß eine Zurückweisung erfolgen. Bevorzugt wird für diesen Fall eine Schranke der Unterschiede in den Trefferraten festgelegt, bei deren Unterschreiten eine Zurückweisung erfolgt, da mit der Überwachung einer Schranke lediglich ein geringer Rechenaufwand verbunden ist.

Ein Vorteil des aufgezeigten Verfahrens besteht darin, daß ein statistisches Ähnlichkeitsmaß eingesetzt wird, welches es erlaubt, aus einer gegebenen Anzahl von verschiedenen Lautmodellen für ähnliche Laute in unterschiedlichen Sprachen dasjenige Lautmodell auszuwählen, welches in seiner Charakteristik alle zur Verfügung stehenden Merkmalsvektoren der jeweiligen Laute am besten beschreibt.

Vorteilhaft wird als Maß für die Auswahl des besten HMM für unterschiedliche Lautmerkmalsvektoren der logarithmische Wahrscheinlichkeitsabstand zwischen den jeweiligen HMM und einem jeden Merkmalsvektor ermittelt. Hierdurch wird ein Maß zur Verfügung gestellt, welches experimentelle Befunde bezüg-

lich der Ähnlichkeit von einzelnen Lautmodellen und deren Erkennungsraten widerspiegelt.

Vorteilhaft wird als Maß für die Beschreibung eines möglichst repräsentativen HMM der arithmetische Mittelwert der logarithmischen Wahrscheinlichkeitsabstände zwischen jedem HMM und den jeweiligen Merkmalsvektoren gebildet, da hierdurch ein symmetrischer Abstandswert erhalten wird.

Vorteilhaft wird das Beschreibungsmaß für die repräsentative Eigenschaft eines HMM zur Beschreibung von Lauten in unterschiedlichen Sprachen dadurch gebildet, daß die erfindungsgemäßen Gleichungen 5 bis 8 angewendet werden, da hierdurch ein geringer Rechenaufwand entsteht.

Vorteilhaft wird für die Anwendung eines Beschreibungsmaßes eine Schrankenbedingung vorgegeben, mit der eine Erkennungsrate des repräsentierenden HMM eingestellt werden kann.

Besonders vorteilhaft wird durch das Verfahren der Speicheraufwand für eine Sprachbibliothek reduziert, da ein Modell für mehrere Sprachen verwendet werden kann. Ebenfalls wird der Portierungsaufwand von einer Sprache in die andere minimiert, was einen reduzierten Zeitaufwand für die Portierung bedingt, der sich durch die On-Line-Adaption auch auf Null vermindern kann. Ebenso wird vorteilhaft ein geringerer Rechenaufwand bei der Viterbi-Suche ermöglicht, da beispielsweise bei mehrsprachigen Eingabesystemen weniger Modelle überprüft werden müssen.

Besonders vorteilhaft werden besondere HMM zur Verwendung in mehrsprachigen Spracherkennungssystemen eingesetzt. Durch diese Vorgehensweise können für Laute in mehreren Sprachen zu Polyphonem-Modellen zusammengefaßte HMM eingesetzt werden. Bei denen Überlappungsbereiche der verwendeten Standardwahrscheinlichkeitsdichteverteilungen bei den unterschiedlichen Modellen untersucht wurden. Zur Beschreibung des Poly-

phonem-Modelles kann eine beliebige Anzahl von identisch bei den unterschiedlichen Modellen verwendeten Standardwahrscheinlichkeitsdichteverteilungen herangezogen werden. Vorteilhaft können auch mehrere Standardverteilungen aus unterschiedlichen Sprachmodellen verwendet werden, ohne daß die hierdurch bewirkte Verwischung der einzelnen Sprachcharakteristika zu einer signifikant niedrigeren Erkennungsrate beim Einsatz dieses Modells führen würde. Als besonders vorteilhaft hat sich hier der Abstandsschwellenwert fünf zwischen ähnlichen Standardwahrscheinlichkeitsverteilungsdichten bewährt.

Besonders vorteilhaft werden beim Einsatz des Verfahrens mit drei Zuständen aus Anlaut, Mittellaut und Ablaut modellierte HMM verwendet, da hierdurch eine hinreichende Genauigkeit bei der Beschreibung der Laute erzielt wird und der Rechenaufwand bei der Erkennung und On-Line-Adaption in einem Spracherkenner gering bleibt.

Im folgenden werden Ausführungsbeispiele der Erfindung anhand von Figuren weiter erläutert.

Figur 1 zeigt ein Blockdiagramm eines Spracherkennungsverfahrens mit Codebuchadaption.

Figur 2 zeigt dabei den Aufbau eines einzigen Multilingualen Phonemes.

In Figur 1 wird in Form eines Blockdiagramms schematisch erläutert, welche einzelnen Bearbeitungsschritte das Verfahren, bzw. ein Spracherkennungssystem, das nach dem Verfahren arbeitet, erfordert.
In einer Erkennungsstufe ERKE des Spracherkennungssystems wird Sprache SPRA erkannt und als Wort WO ausgegeben. Es können auch Untereinheiten von Worten durch hidden-Markov-Modelle HMM modelliert worden sein und als Worte WO ausgegeben werden. In einem Lexikon LEX des Spracherkennungssystems sind beispielsweise als vom Hersteller vorgegebene Wort-

strings ANW für die Anwendung in Form von ASCII-Zeichen abgelegt. In einem HMM-Codebuch CB sind zuvor trainierte und mit dem Lexikon LEX ausgelieferte Parameter für hidden-Markov-Lautmodelle abgelegt. Für eine mehrsprachige Anwendung des Spracherkennungssystems, kann das Lexikon auch HMM enthalten, die speziell für eine Mehrsprachenanwendung bereitgestellt, bzw. gebildet werden. Anhand des Lexikons LEX und des HMM-Codebuches CB führt der Spracherkenner ERKE die Erkennung von Worten aus Sprachsignalen SPRA durch. Zur Anpassung des Spracherkennungssystems an eine spezifische Anwendung, kann das Lexikon LEX beispielsweise vom Anwender durch anwendungsspezifische Wortstrings ANW abgeändert werden. Hierzu können gegebenenfalls auch Wortstrings in einer Fremdsprache eingegeben werden. Fremdsprache bedeutet in diesem Zusammenhang, daß die Sprache bei der Bereitstellung des Codebuches nicht berücksichtigt wurde. Gemäß dem Verfahren wird nach Erkennung eines speziellen Wortes oder einer Erkennungseinheit WO, einem Adaptionsbaustein ADAP über eine Verbindungsleitung 100 mitgeteilt, welches dieser Worte erkannt wurde und welche Segmente damit verbunden sind. Anschließend erfolgt bevorzugt eine Anpassung, der mit dem erkannten Wort verbundenen Parameter der hidden-Markov-Modelle an den Merkmalsvektor, welcher aus dem Sprachsignal abgeleitet wurde. Im Adaptionsbaustein ADAP kann beispielsweise eine bevorzugt auszuführende Adaptionsstrategie zur Anpassung der hidden-Markov-Modelle festgelegt sein. In einer Anpassungsvariante können beispielsweise Worte mit unsicheren Trefferraten für die einzelnen Hypothesen nach der Viterbi-Suche, ausgelassen werden. Da erfindungsgemäß neue Koartikulationen gelernt werden sollen, können bevorzugt lediglich nur solche Merkmalsvektoren zu Anpassung ausgewählt werden, welche speziell den neu zu lernenden Koartikulationssegmenten zugeordnet werden. Fallweise kann es jedoch günstiger sein alle zur Verfügung stehenden Merkmalsvektoren zur Anpassung auszuwählen, um sicherzustellen, daß auch solche Koartikulationen von der Anpassung erfaßt werden, welche über ein Diphon hinaus reichen.

Der Schwerpunktsvektor der zugrundeliegenden hidden-Markov-Modelle wird an den Merkmalsvektor angepaßt, indem beispielsweise Komponentenweise eine Mittelwertbildung durchgeführt wird und diese Mittelwertbildung zu einer Verschiebung des im Codebuch CB gespeicherten Merkmalsvektors führt. Hierzu werden die jeweiligen Mittelwerte mit einem Anpassungsfaktor, der hier als Lernschrittweite fungiert, multipliziert, so daß ein neuer Schwerpunktsvektor des im Lexikon gespeicherten hidden-Markov-Modelles bzw. der gespeicherten hidden-Markov-Modelle entsteht. Dieser adaptierte Schwerpunktsvektor fungiert in Zukunft als Ausgangsgröße bei der Erkennung von Sprachsignalen im Spracherkenner ERKE.

Die Grundidee besteht dabei darin, daß das System während der Anwendung beim Auftreten eines veränderten und vom Anwender vorgegebenen Lexikons automatisch nachtrainiert bzw. nachadaptiert wird. Beispielsweise wird eine solche Veränderung festgestellt, indem ins Lexikon LEX eingegebene Wortstrings ANW mit dem Lexikoninhalt verglichen werden. Auf diese Weise können auch Wortstrings in einer Fremdsprache einfach identifiziert werden, um ggf. spezielle Mehrsprachen-HMM heranzuziehen. Vorzugsweise erfolgt gleichzeitig mit der Eingabe des Wortstrings in das Lexikon eine erste Eingabe des Wortes als Sprache SPRA, um eine erste Zuordnung zwischen den im Codebuch CB vorhandenen HMM und dem neu zu erkennenden Wort herzustellen. Dieses adaptive Nachtuning der Erkennungsparameter erfolgt gemäß der Erfindung anhand von Sprachdaten, welche während der Bedienung des Systems anfallen. Bevorzugt wird die Adaption dabei bei jeder Änderung nachgeführt, ohne daß während der Entwicklungsphase des Spracherkennungssystems das jeweilige Vokabular für die Erstellung des Lexikons LEX bekannt sein muß. Gegenüber dem Stand der Technik weist das erfindungsgemäße Verfahren den Vorteil auf, daß es On-Line abläuft, ohne das ein Satz spezieller Sprachproben für das Training benötigt wird. Hierdurch ergibt sich ebenfalls die Möglichkeit Mehrsprachen-HMM On-Line an eine Fremdsprache anzupassen. Gemäß dem Adaptionsver-

fahren erfolgt die Anpassung dabei bevorzugt unüberwacht im Hintergrund des Systems, wozu es seine eigenen Ergebnisse zur Adaption während der Anwendung verwendet. Die dabei benötigten Rechenschritte sind relativ einfach zu implementieren und erfordern eine geringe Rechenleistung.

Der grundlegende Gedanke besteht dabei darin, daß die Spracherkennung auf HMM basiert. Beim Training solcher Modelle werden insbesondere die Parameter zur Berechnung der Emissionswahrscheinlichkeiten bestimmt. Die zur Berechnung benötigten Wahrscheinlichkeitsdichten werden durch Standardverteilungen, wie z. B. Gauß-, oder Laplace-Verteilungen angenähert. Wichtigster Parameter für diese Approximation ist dabei der Mittelpunktsvektor, bzw. der Schwerpunktsvektor der jeweiligen Verteilungen. Diese Parameter sind im Codebuch gespeichert. Während der Spracherkennung liegt bei der Erkennung mit dem sogenannten Viterbi-Algorithmus nach der Klassifizierung, eine Zuweisung einzelner Spracheinheiten, welche durch Merkmalsvektoren repräsentiert werden, zu bestimmten Erkennungssegmenten und den entsprechenden Wahrscheinlichkeitsdichteverteilungen vor. Nach dem aufgezeigten Verfahren erfolgt der eigentliche Adaptionsschritt bevorzugt durch eine Neuberechnung der Mittelpunkte der betroffenen Wahrscheinlichkeitsdichteverteilungen unter Benutzung der in der Anwendung angefallenen Merkmalsvektoren. Besonders vorteilhaft wird dabei die Adaption nach jeder abgeschlossenen Äußerung ausgeführt, sobald der Viterbi-Pfad mit der eindeutigen Zuordnung von Merkmalsvektor zu Wahrscheinlichkeitsdichteverteilung vorliegt.

Ein Problem welches der Erfindung löst besteht dabei darin, daß das Training eines großen wortschatzunabhängigen hidden-Markov-Modelles, welches mit allen Erfordernissen aus allen denkbaren praktischen Anwendungen fertig wird, nicht möglich ist [1]. An praktische Anwendungen sind dabei besonders strenge Anforderungen zu stellen. Adaptionsverfahren zur Spracherkennung sollten dabei

- wenig rechenaufwendig und einfach zu implementieren
- unüberwacht
- sprecherunabhängig
- On-Line arbeiten und im voraus kein vorheriges Adaptionsset erfordern. Besonders soll für die Anwendung in dem erfindungsgemäßen Verfahren ein HMM-Codebuch als Saatmodell eingesetzt werden, welches wortschatzunabhängig trainiert wurde, so daß es keine Merkmale und Bevorzugungen von irgendwelchen speziellen Erkennungseinheiten aufweist. Beispielsweise können die zugrundeliegenden HMM als monophone Modelle trainiert sein, jedoch können auch hidden-Markov-Modelle mit verbundenen Diphonen eingesetzt werden. Bei der Erprobung des erfindungsgemäßen Verfahrens wurden als Saatmodell hidden-Markov-Modelle verwendet, welche monophon trainiert wurden. Die Strategie bei der Anpassung des Codebuches nach dem Verfahren besteht dabei beispielsweise darin, sein allgemeines monophones Saatmodell, beispielsweise auch für Mehrsprachen HMM, als Ausgangsbasis zu verwenden und sein phonemisches Inventar zur Erstellung eines arbeitsfähigen Diphon-Modelles zu verwenden, wann immer das Lexikon verändert wird und ein neues kontextabhängiges Segment für geänderte Betriebserfordernisse erstellt werden muß. Dabei wird das jeweilige Modell bevorzugt während des Erkennungsprozesses On-Line adaptiert. Hierzu werden bevorzugt folgende Schritte ausgeführt:
- Zunächst wird das Lexikon LEX untersucht um herauszufinden, welche kontextabhängigen Segmente benötigt werden.
- Falls ein auftauchendes Segment bis dahin unbekannt war werden die korrespondierenden kontextunabhängigen Segmentverteilungen vom allgemeinen Modell in das neue Modell des Arbeitswörterbuches kopiert.
- Erkennung von eingehenden Sprachäußerungen.
- Fallweise Zurückweisung von unsicheren Erkennungsergebnissen wenn das gewünscht wird.
- On-Line-Training des Arbeitscodebuches mit der beschriebenen Anpassungsformel auf Basis der eingehenden Sprachdaten.

Zur Adaption der Modelle, wird der Schwerpunktsvektor der erkannten hidden-Markov-Modelle an den Merkmalsvektor des ein-

gehenden Sprachsignales angepaßt. Dabei wird bevorzugt gemäß einer Lernrate eine Verschiebung des im Lexikon gespeicherten Schwerpunktsvektors in Richtung des erkannten Merkmalsvektors der entsprechenden Lautäußerungen durchgeführt. Dabei wird davon ausgegangen, daß die relevanten Unterschiede zwischen den aufgabenspezifischen Versionen des Lexikons hauptsächlich die Parameter der Wahrscheinlichkeitsdichteverteilung der hidden-Markov-Modelle angehen, wobei insbesondere der Ort der Mittelwerte im akustischen Raum betroffen ist.

Von einer Merkmalsextraktionsstufe eines Spracherkennungssystems wird dabei eine eingehende Äußerung bevorzugt in eine Serie von Merkmalsvektoren transformiert:

$$X = \left\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_T\right\} \qquad (1)$$

Unter Verwendung des Viterbi-Algorithmus wird dann beispielsweise jeder einzelne Merkmalsvektor $\vec{x}_t$ mit t = 1, 2, ... T einem Zustand Θ^i_t des besten hidden-Markov-Modelles i nach der Erkennung zugewiesen. Für den Fall, daß multimodale Laplace-Verteilungen für die Modellierungen der hidden-Markov-Modelle und der Zustands-Emissionswahrscheinlichkeiten verwendet werden, läßt sich die korrespondierende Wahrscheinlichkeitsdichteverteilung des S-ten Zustandes eines hidden-Markov-Modelles wie folgt approximieren

$$b^i_s(\vec{x}) = \sum_{m=1}^{M^i_s} c^i_{s,m} e^{\frac{\sqrt{2}}{\sigma}\left\|\vec{x}-\vec{\mu}^i_{s,m,t}\right\|} \qquad (2)$$

Dabei sind $M^i_s, c^i_{s,m}$ und σ Konstanten, welche bevorzugt während des Trainings bestimmt werden. Bei einer gegebenen Zuordnung eines erkannten Merkmalsvektors und eines Zustandes wird dann bevorzugt der Mittelwert $\vec{\mu}^i_{s,m,t}$ bestimmt, welcher am nächsten am Merkmalsvektor $\vec{x}_t$ liegt, wobei der City-Blockabstand (2*) als Maß benutzt wird und n die Komponente eines Vektors bezeichnet.

$$\|\vec{x}-\vec{\mu}\| = \sum_n |x_n - \mu_n| \qquad (2^*)$$

Der am Nächsten liegende mittlere Abstand wird dabei gemäß

$$\vec{\mu}^i_{s,m,t+1} = (1-\alpha)\vec{\mu}^i_{s,m,t} + \alpha\vec{x}_t \qquad (3)$$

aktualisiert. Gleichung 3 kann man sich geometrisch wie folgt interpretiert vorstellen. Der aktualisierte Schwerpunktsvektor $\vec{\mu}^i_{s,m,t+1}$ liegt auf einer Geraden, welche durch den alten Mittelpunktsvektor $\vec{\mu}^i_{s,m,t}$ und den aktuellen Merkmalsvektor $\vec{x}_t$ geht. Der Parameter α wird dabei als Adaptionsfaktor oder als Lernrate verwendet. Für den speziellen Fall, daß $\alpha = 0$ ist, wird keine Adaption durchgeführt, während für $\alpha = 1$ der aktuelle Schwerpunktsvektor dem aktuellen Merkmalsvektor entspricht.

In allen Anwendungen von Dialogsystemen können ebenfalls Erkennungsfehler auftreten. Die Ursachen bestehen dabei beispielsweise in einem falschen Eingabewort durch den Benutzer oder einfach in einer falschen Klassifizierung durch die Erkennungseinheit. Für den Fall daß ein solcher Fehler auftritt, sollte beispielsweise ein Dialogmanager den Benutzer zu einer erneuten Eingabe auffordern. Vereinfacht kann aber auch lediglich eine besonders gute Äußerung ausgewählt werden. Hierzu wird beispielsweise eine relativ einfache statistische Zurückweisungsstrategie verwendet. Die Trefferrate S_0 der besten und die Trefferrate S_1 der zweitbesten Lauthypothese nach der n-Bestensuche im Viterbi-Algorithmus wird dabei untersucht.

$$\text{rejectionflag} = \begin{cases} 1 & \text{if} \quad (s_1 - s_0) \le r_{thresh} \\ 0 & \text{else} \end{cases} \qquad (4)$$

Falls die Zurückweisungsmarke rejectionflag 1 beträgt, so wird die korrespondierende Äußerung bevorzugt durch den Adaptionsalgorithmus unterdrückt. Bevorzugt wird die Schranke für die Zurückweisung dadurch bestimmt, daß empirisch S_{mean} d. h.

der Mittelwert aller Trefferraten pro Wort von eingehenden Äußerungen ermittelt wird. Bevorzugt ergibt sich aus Experimenten der Grenzwert $R_{thresh} = 0,005\ S_{mean}$. Mit diesem Grenzwert wird eine Zurückweisungsrate von falschen Äußerungen von 61,2 % erreicht und die Adaption kann dabei mit sicherer klassifizierten Daten durchgeführt werden, als dies der Fall wäre, wenn keine Zurückweisung durchgeführt würde. Bei experimentellen Überprüfung des Verfahrens wurde von 2000 Test-Äußerungen eines geänderten Vokabulars für die Adaption ausgegangen. Bezüglich des Adaptionsfaktors α und seiner Dimensionierung wurde dabei festgestellt, daß bereits kleine Werte von α, d. h. 0,025 die Fehlerrate bei der Erkennung signifikant verringern. Ein breites Optimum von α wurde dabei zwischen 0,05 und 0,01 festgestellt. Dabei wurde bei einem Optimum von 0,075 eine Verbesserung der Fehlerrate von 34,5 % bei der Erkennung von Worten erzielt. Das bedeutet, daß sich durch das erfindungsgemäße Verfahren Erkennungseinheiten mit dem angepaßten Codebuch CB um 34,5 % besser erkennen lassen, als dies ohne seine Anwendung der Fall wäre.

Für den Fall daß eine Zurückweisung wie beschrieben durchgeführt wird, ergibt sich ein verbesserter Wert des Anpassungsfaktors α zu 0,125. Dieser verbesserte Wert von α führt zu einer Reduktion der Fehlerrate von 40,1 % bei dem experimentell verwendeten Wortschatz. Der höhere Faktor von α läßt sich dadurch erklären, daß durch die Zurückweisung von falschen Daten ein besserer Merkmalsvektor für die Anpassung des HMM-Codebuches vorliegt und daß damit eine höhere Lernschrittweite gewählt werden kann. Die experimentellen Befunde haben auch gezeigt, daß mit dem adaptiven Verfahren nahezu dieselbe Erkennungsrate erreicht wird, wie dies für ein spezielles Modell für den entsprechenden Anwendungsfall erzielt würde. Die Erkennungsrate lag dabei nur 0,7 % unter der des Speziallexikons.

Figur 2 zeigt den Aufbau eines einzigen Multilingualen Phonemes. In diesem Fall ist es das Phonem M das dargestellt wird. Die Zahl der Wahrscheinlichkeitsdichten und die Erkennungsrate für dieses Phonem sind in Tabelle 1 angegeben:

Thr.	#densit(a,b,c).	Engl.[%]	Germ.[%]	Span.[%]
0	341(0 0 341)	46.7	44.7	59.4
2	334(0 14 327)	45.0	46.4	57.5
3	303(27 34 280)	48.0	45.8	57.5
4	227(106 57 187)	50.9	44.1	58.7
5	116(221, 48,72)	49.3	43.1	57.0
6	61(285, 22, 34)	41.2	38.6	50.4

In Figur 2 ist der Anlaut L, der Mittellaut M und der Ablaut R des Phonem-Modelles dargestellt. Für die unterschiedlichen Sprachen Englisch EN, Deutsch DE und Spanisch SP sind die Schwerpunkte der Wahrscheinlichkeitdichteverteilungen der einzelnen verwendeten Standardwahrscheinlichkeitdichten eingetragen und als WD gekennzeichnet. Hier ist beispielsweise ein HMM aus drei Teilzuständen dargestellt. Die Erfindung soll jedoch nicht lediglich auf Anwendung solcher HMM beschränkt werden, obwohl diese unter Berücksichtigung des Kriteriums, das ein minimaler Rechenaufwand der Erkennung durchgeführt werden soll ein gewisses Optimum darstellen. Die Erfindung kann ebenso auf HMM angewendet werden, die eine andere Anzahl von Zuständen aufweisen. Durch die Erfindung soll insbesondere erreicht werden, daß der Portierungsaufwand bei der Portierung von Spracherkennungssystemen in eine andere Sprache reduziert, bzw. vermieden wird und daß die verwendeten Rechenresourcen durch Reduktion der zugrundeliegenden Parameter möglichst gering gehalten werden. Beispielsweise können durch derartige Spracherkennungssysteme begrenzte Hardwareerfordernisse besser erfüllt werden, insbesondere wenn ein- und dasselbe Spracherkennungssystem für Mehrsprachenanwendung in einem Gerät zur Verfügung gestellt werden soll.

Zunächst sollte um das Ziel zu erreichen, die Ähnlichkeiten von Lauten in unterschiedlichen Sprachen auszuschöpfen und beim Modellieren zu berücksichtigen, beachtet werden, daß sich die Phoneme in verschiedenen Sprachen unterscheiden können. Die Gründe hierfür bestehen vor allen Dingen in:

- Unterschiedlichen phonetischen Kontexten, wegen der unterschiedlichen Phonemsätze in den verschiedenen Sprachen;
- unterschiedlichen Sprechweisen;
- verschiedenen prosodischen Merkmalen;
- unterschiedlichen allophonischen Variationen.

Ein besonders wichtiger Aspekt, welcher dabei zu berücksichtigen ist, besteht im Prinzip der genügenden wahrnehmungstechnischen Unterscheidbarkeit der Phoneme [7]. Dies bedeutet, daß einzelne Laute in verschiedenen Sprachen akustisch unterscheidbar gehalten werden, so daß es für den einzelnen Zuhörer leichter ist sie voneinander zu separieren. Da aber jede einzelne Sprache einen unterschiedlichen Phonemschatz hat, werden die Grenzen zwischen zwei ähnlichen Phonemen in jeder einzelnen Sprache sprachspezifisch festgelegt. Aus diesen Gründen hat die Ausprägung eines bestimmten Lautes eine sprachspezifische Komponente.

Bevorzugt werden die zugrundeliegenden Phoneme mittels kontinuierlichen dichten hidden-Markov-Modellen (CD-HMM) modelliert [5]. Als Dichtefunktionen werden häufig Laplace-Mischungen benutzt. Dabei besteht jedes einzelne Phonem aus drei Zuständen von links nach rechts gerichteten HMM. Die akustischen Merkmalsvektoren bestehen dabei beispielsweise aus 24 mel-skalierten cepstral, 12 delta cepstral, 12 delta delta cepstral, Energie, delta-Energie und delta delta - Energie-Koeffizienten. Beispielsweise wird als Länge des Untersuchungszeitfensters 25 ms gewählt, wobei die Rahmenabstände 10 ms zwischen den einzelnen Rahmen betragen. Aus Gründen der begrenzten Größe des Sprachkorpus werden bevorzugt lediglich kontextunabhängige generierte Phoneme ange-

wandt. Als besonders repräsentatives Phoneminventar wird jenes aus [6] gewählt.

Die Idee des Verfahrens besteht darin, daß zum einen ein zur Verfügung gestelltes Ähnlichkeitsmaß verwendet wird, um aus standardmäßig verfügbaren Sprachphonembibliotheken für unterschiedliche Sprachen jenes HMM auswählen zu können, welches den Merkmalsvektoren, die aus den unterschiedlichen Lautmodellen der unterschiedlichen Sprachen abgeleitet werden, am nächsten kommt. Hierdurch ist es möglich, die Ähnlichkeiten zweier Phonem-Modelle zu ermitteln und über dieses Ähnlichkeitenmaß basierend auf der Differenz der Log-Likelihood-Werte zwischen den Lautrealisierungen und Lautmodellen eine Aussage zu treffen, ob es sich lohnt, einen Laut für mehrere Sprachen gemeinsam zu modellieren, bzw. ein betreffendes schon bestehendes HMM für die Modellierung des Lautes in mehreren Sprachen zu verwenden. Hierdurch wird die Zahl der Parameter, welche bei der Spracherkennung und Adaption der Modelle zu berücksichtigen sind reduziert, indem die Zahl der zu untersuchenden HMM reduziert wird.

Ein weiterer Lösungsansatz besteht darin, ein spezielles zur Modellierung eines Lautes in mehreren Sprachen zu erstelltes Polyphonem-Modell zu verwenden. Zu dessen Erzeugung werden zunächst beispielsweise drei Lautsegmente, in Form eines Anlautes, Mittellautes und Ablautes gebildet, deren Zustände aus mehreren Wahrscheinlichkeitdichtefunktionen den sogenannten Mischverteilungsdichten mit den dazugehörigen Dichten bestehen. Diese Dichten der über verschiedenen Sprachen ähnlichen Lautsegmente werden dann zu einem multilingualen Codebuch zusammengefaßt. Somit teilen sich Lautsegmente verschiedener Sprachen die gleichen Dichten. Während das Codebuch für mehrere Sprachen gleichzeitig benutzt werden kann, werden beispielsweise die Gewichte, mit denen die Dichten gewichtet werden für jede Sprache getrennt ermittelt und bereitgestellt.

Zur Bildung eines geeigneten Ähnlichkeitsmaßes werden dabei bevorzugt HMM mit drei Zuständen herangezogen. Das Abstands- oder Ähnlichkeitsmaß kann dabei benutzt werden um mehrere Phonem-Modelle zu einem multilingualen Phonem-Modell zusammenzufassen oder diese auf geeignete Weise zu ersetzen. Hierdurch kann ein multilingualer Phonemschatz bereitgestellt werden. Bevorzugt wird zur Messung des Abstandes bzw. zur Bestimmung der Ähnlichkeit von zwei Phonem-Modellen des selben Lautes aus unterschiedlichen Sprachen eine Meßgröße verwendet, welche auf der relativen Entropie basiert [3]. Während des Trainings werden dabei die Parameter der gemischten Laplacedichteverteilungen der Phonem-Modelle bestimmt. Weiterhin wird für jedes Phonem ein Satz von Phonemtokens X als Merkmalsvektor aus einem Test- oder Entwicklungssprachkorpus extrahiert. Diese Phoneme können dabei durch ihr international genormtes phonetisches Etikett markiert sein. Demnach werden zwei Phonem-Modelle λ_i und λ_j und ihre zugehörigen Phonemtoken X_i und X_j zur Bestimmung des Ähnlichkeitsmaßes zwischen diesen unterschiedlichen Phonemen wie folgt behandelt.

$$d(\lambda_i,\lambda_j) = \log p(X_i|\lambda_i) - \log p(X_i|\lambda_j) \qquad (5)$$

Dieses Abstandsmaß kann als Log-Likelihood-Abstand angesehen werden, welcher darstellt wie gut zwei verschiedene Modelle zu dem selben Merkmalsvektor X_I passen. Demgemäß wird der Abstand zwischen den beiden Modellen λ_i und λ_j gemäß:

$$d(\lambda_j,\lambda_i) = \log p(X_j|\lambda_j) - \log p(X_j|\lambda_i) \qquad (6)$$

bestimmt. Um einen symmetrischen Abstand zwischen diesen beiden Phonem-Modellen zu erhalten, wird dieser bevorzugt gemäß

$$d(\lambda_j;\lambda_i) = \frac{1}{2}\left(d(\lambda_i,\lambda_j) + d(\lambda_j;\lambda_i)\right) \qquad (7)$$

bestimmt. Anhand von experimentellen Befunden konnte festgestellt werden, daß sich durchaus einige Phonem-Modelle aus anderen Sprachen besser für die Verwendung in einem deutschen Spracherkennungssystem eignen, als ein deutsches Phonem-Mo-

dell. Beispielsweise gilt dies für die Phoneme k, p und N. Für diese Phoneme eignet sich das englische Phonem-Modell besser als das deutsche. Während beispielsweise ein großer Unterschied zwischen dem deutschen und dem englischen Modell über den Umlaut aU beobachtet wurde, was bedeutet, daß für beide Laute ein unterschiedliches Symbol im multilingualen Phonemschatz eingeführt werden sollte. Andererseits konnte für den Umlaut aI im deutschen und im englischen eine große Ähnlichkeit festgestellt werden, das bedeutet, daß lediglich ein Phonem-Modell für beide Sprachen gleich gut Verwendung finden kann. Ausgehend davon sollte für jedes Symbol eines multilingualen Phonemschatzes ein separates statistisches Modell erzeugt werden. In [8] wurden Polyphoneme als solche Phoneme bezeichnet, die ähnlich genug sind, um in verschiedenen Sprachen als ein einziges Phonem modelliert zu werden. Ein Nachteil dieser Vorgehensweise besteht darin, daß für die sprachspezifische Erkennung der vollständige akustische Raum des Polyphonems verwendet wird. Ziel ist es jedoch, die sprachabhängigen und die sprachspezifischen akustischen Eigenschaften eines multilingualen Modells zu kombinieren. Gemäß der Erfindung sollen in einem Polyphonem-Modell solche Bereiche des akustischen Raumes eingegrenzt sein , in denen sich die verwendeten Wahrscheinlichkeitdichten der einzelnen Phoneme überlappen. Dazu wird z.B. eine gruppierende Verdichtungstechnik (agglommerative density clustering technique) eingesetzt, um gleiche oder ähnliche Ausprägungen eines Phonems zu reduzieren. Besonders wichtig ist es dabei zu beachten, daß lediglich die Dichten der korrespondierenden Zustände der einzelnen HMM in den Phonemen zusammengefaßt werden dürfen.

In Figur 2 ist dabei zu erkennen, daß die jeweiligen Dichten für die einzelnen Zustände L, M und R in den eingegrenzten Regionen enthalten sind. Während identische Dichten über die einzelnen Sprachen EN, DE, und SP verteilt sind, variieren die Mischungsgewichte sprachabhängig. Zu berücksichtigen ist jedoch, daß spezifische Ausprägungen eines Phonems in ver-

schiedenen Sprachen in unterschiedlicher Häufigkeit auftreten.

Die Zusammenfassung der unterschiedlichen Wahrscheinlichkeitsdichten kann dabei mit einem unterschiedlichen Abstandsschwellenwert für die Wahrscheinlichkeitsdichten bei der Dichtehäufung (density clustering) durchgeführt werden. Beispielsweise reduziert sich mit einem Abstandsschwellenwert von fünf die Zahl der verwendeten Dichten um einen Faktor 3 gegenüber dem Ausgangszustand, ohne daß damit eine entscheidende Verschlechterung bei der Spracherkennungsrate einhergeht. In diesem Beispiel wurden 221, 48 und 72 von den ursprünglichen 341 Ausgangsdichten für jeweils die Polyphonem-Region, die Zweisprachen-Region und die Einsprachen-Region zusammengefaßt. In Figur 2 ist eine solche Polyphonemregion als Schnittmenge der Kreise für die einzelnen Sprachen dargestellt. Beim Mittellaut M des dargestellten HMM ist eine Wahrscheinlichkeitsdichte in einer solchen Region als WDP bezeichnet. Die Erkennungsraten für ein komplettes multilinguales Spracherkennungssystem sind dabei in Spalte 4 und 5 der Tabelle 2 als ML1 und ML2 angegeben:

Language	#Tokens	LDP[%]	ML1[%]	ML2[%]
English	21191	39.0	37.3	37.0
German	9430	40.0	34.7	37.7
Spanish	9525	53.9	46.0	51.6
Total	40146	42.8	38.8	40.8

Während bei der ersten Untersuchung ML1 die konventionelle Polyphonem-Definition aus [8] verwendet wurde, was bedeutet, daß der komplette akustische Bereich des Polyphonem-Modells bestehend aus der äußeren Kontur der Sprachbereiche in Figur 2, für die Erkennung verwendet wurde, benutzt die beschriebene Methode lediglich einen Teilbereich daraus. Durch Verwendung der teilweisen Überlappung der einzelnen Sprachbereiche für die Modellierung des Polyphonem-Modells, ist beispiels-

weise eine Verbesserung von 2 % erzielbar, wie dies in Tabelle 2 in der Spalte für ML2 dargestellt ist.

Durch die Kombination multilingualer Lautmodelle mit dem automatischen On-Line- Adaptionsverfahren, läßt sich die Erkennungsgenauigkeit der Modelle verbessern. Durch eine unüberwacht Adaption werden sie an das neue Anwendungsvokabular, bzw. die neue Sprache angepaßt. Hierzu müssen vorteilhaft keine zusätzlichen Daten für die neue Sprache gesammelt werden. Falls universelle Lautmodelle eingesetzt werden, kann das Verfahren für beliebige Sprachen verwendet werden. Beispielsweise kann mit multilingualen Lautmodellen aus Daten der Sprachen Deutsch, Amerikanisch und Spanisch durch On-Line-Adaption ein Einzelworterkennungssystem für slowenische Ziffern optimiert werden. Hierzu wird dem Erkenner beim Erkennungsvorgang lediglich slowenisches Datenmaterial zugeführt.

Literatur:

[1] Hon H.W., Lee K.F., „On Vocabulary-Independent Speech Modeling", Proc. IEEE Intern. Conf. on Acoustics, Speech, and Signal Processing, Albuquerque NM, 1990;

[2] Lee C.H., Gauvain J.L., „Speaker Adaptation Based on MAP Estimation of HMM Parameters", Proc. IEEE Intern. Conf. on Acoustics, Speech, and Signal Processing, Minneapolis MN, 1993;

[3] V. Digalakis A. Sankar, F. Beaufays.: „Training Data Clustering For Improved Speech Recognition.", In Proc. EUROSPEECH '95, pages 503 - 506, Madrid, 1995;

[4] P. Dalsgaard and O. Andersen.: „Identification of Mono- and Poly-phonemes using acousitc-phonetic Features derived by a self-organising Neural Network.", In Proc. ICSLP '92, pages 547 - 550, Banff, 1992;

[5] A. Hauenstein and E. Marschall.: „Methods for Improved Speech Recognition Over the Telephone Lines.", In Proc. ICASSP '95, pages 425 - 428, Detroit, 1995

[6] J. L. Hieronymus.: „ASCII Phonetic Symbols for the World's Languages: Worldbet.", preprint, 1993;

[7] P. Ladefoged: „A Course in Phonetics.", Harcourt Brace Jovanovich, San Diego, 1993;

[8] P. Dalsgaard O. Andersen and W. Barry.: „Data-driven Identification of Poly- and Mono-phonemes for four European Languages.", In Proc. EUROSPEECH '93, pages 759 - 762, Berlin, 1993;

[9] A. Cole Y.K. Muthusamy and B.T. Oshika.: „The OGI Multilanguage Telephone Speech Corpus.", In Proc. IC-SLP '92, pages 895 - 898, Banff, 1992;

[10] B. Wheatley, K. Kondo, W.Anderson, Y. Muthusamy: „An Evaluation Of Cross-Language Adaption For Rapid HMM Development In A New Language", In Proc. ICASSPP ' Adelaide,1994, pages 237 - 240;

Patentansprüche

1. Verfahren zur Echtzeit-Anpassung eines hidden-Markov-Lautmodelles im Codebuch eines Spracherkennungssystems an eine Wortschatzänderung im verwendeten phonetischen Lexikon,
   a) bei dem zu erkennende hidden-Markov-Lautmodelle mindestens über einen ersten Mittelwertsvektor ihrer Wahrscheinlichkeitsverteilungen im Codebuch (CB) verfügbar gehalten werden,
   b) bei dem die Spracherkennung (ERKE) in üblicher Weise durch Extraktion von Merkmalsvektoren aus einem Sprachsignal (SPRA) und Zuordnung der Merkmalsvektoren zu Wahrscheinlichkeitsverteilungen von hidden-Markov-Lautmodellen aus dem Codebuch (CB) durchgeführt wird,
   c) und bei dem für mindestens eine erkannte Lautäußerung (WO) der Wortschatzänderung unmittelbar nach deren Erkennung die Lage des ersten Schwerpunktsvektors mindestens eines zugehörigen hidden-Markov-Lautmodelles an die Lage des zugeordneten Merkmalsvektors über einen festgelegten Anpassungsfaktor maßstäblich angepaßt (ADAP, CB, 100) und der angepaßte Mittelwertsvektor im Codebuch (CB) als erster Mittelwertsvektor abgelegt wird.

2. Verfahren nach Anspruch 1, bei dem die Anpassung der Vektorlage durch komponentenweise Mittelwertbildung und Multiplikation mit einem konstanten Anpassungsfaktor durchgeführt wird.

3. Verfahren nach einem der vorangehenden Ansprüche, bei dem die Zuordnung der Merkmalsvektoren zu den entsprechenden hidden-Markov-Lautmodellen mit Hilfe des Viterbi-Algorithmus durchgeführt wird.

4. Verfahren nach einem der vorangehenden Ansprüche,
   a) bei dem für die Spracherkennung eine Folge von Merkmalsvektoren der Form

$$X = \{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_T\} \quad (1)$$

aufgenommen wird,

b) bei dem anzupassende und zu erkennende hidden-Markov-Lautmodelle mindestens je über einen ersten Schwerpunksvektor ihrer Laplace-Wahrscheinlichkeitsverteilungen der Form

$$b_s^i(\vec{x}) = \sum_{m=1}^{M_s^i} c_{s,m}^i e^{-\frac{\sqrt{2}}{\sigma}\left\|\vec{x} - \vec{\mu}_{s,m,t}^i\right\|} \quad (2)$$

mit beim Training bestimmten Konstanten M_s^i $c_{s,m}^i$ σ verfügbar gehalten werden,

c) und bei dem für mindestens eine erkannte Lautäußerung nach deren Erkennung die Lage des ersten Schwerpunktsvektors mindestens eines zugehörigen hidden-Markov-Lautmodelles an die Lage des betreffenden Merkmalsvektors über

$$\vec{\mu}_{s,m,t+1}^i = (1-\alpha)\vec{\mu}_{s,m,t}^i + \alpha\vec{x}_t \quad (3)$$

angepaßt wird mit $\vec{\mu}_{s,m,t+1}^i$ als Komponente des neuen Schwerpunktsvektors und α als Anpassungsfaktor.

5. Verfahren nach einem der vorangehenden Ansprüche, bei dem eine nicht erkannte Lautäußerung zurückgewiesen und keine Anpassung durchgeführt wird.

6. Verfahren nach Anspruch 3 und 4, bei dem nach der n-Besten-Suche im Viterbi-Algorithmus eine erste Trefferrate für eine erste Lauthypothese und eine zweite Trefferrate für eine zweite Lauthypothese bestimmt wird und die Zurückweisung mindestens in Abhängigkeit des Unterschiedes zwischen diesen beiden Trefferraten erfolgt.

7. Verfahren nach Anspruch 6, bei dem die Zurückweisung erfolgt, falls der Betrag der Differenz zwischen den beiden Trefferraten eine festgesetzte Schranke unterschreitet.

8. Verfahren nach einem der Ansprüche 1-7 zur Anpassung eines wie folgt gebildeten Mehrsprachen hidden-Markov-Lautmodelles:

a) ausgehend von mindestens einem ersten Mekmalsvektor für einen ersten Laut (L,M,R) in einer ersten Sprache (SP,EN,DE) und von mindestens einem zweiten Mekmalsvektor für einen vergleichbar gesprochenen zweiten Laut in mindestens einer zweiten Sprache (DE,SP,EN) und deren zugehörigen ersten und zweiten hidden-Markov-Lautmodellen wird ermittelt, welches der beiden hidden-Markov-Lautmodelle (L,M,R) beide Merkmalsvektoren besser beschreibt;

b) dieses hidden-Markov-Lautmodell (L,M,R) wird für die Modellierung des Lautes in mindestens beiden Sprachen (SP,EN,DE) verwendet.

9. Verfahren nach Anspruch 8, bei dem als Maß für die Beschreibung eines Merkmalsvektors durch ein hidden-Markov-Lautmodell (L,M,R) der logarithmische Wahrscheinlichkeitsabstand als log likelihood distance zwischen jedem hidden-Markov-Lautmodell und mindestens einem Merkmalsvektor gebildet wird, wobei eine kürzerer Abstand eine bessere Beschreibung bedeutet.

10. Verfahren nach Anspruch 9, bei dem als Maß für die Beschreibung der Merkmalsvektoren durch die hidden-Markov-Lautmodelle der arithmetische Mittelwert der logarithmischen Wahrscheinlichkeitsabstände bzw. der log likelihood distances zwischen jedem hidden-Markov-Lautmodell (L,M,R) und jedem jeweiligen Merkmalsvektor gebildet wird, wobei eine kürzerer Abstand eine bessere Beschreibung bedeutet.

11. Verfahren nach Anspruch 10, bei dem das erste hidden-Markov-Lautmodell (L,M,R) von einem Phonem λ_i und das zweite hidden-Markov-Lautmodell von einem Phonem λ_j verwendet wird und bei dem als erste und zweite Merkmalsvektoren X_i und X_j verwendet werden, wobei der logarithmische Wahrscheinlichkeitsabstand zum ersten Merkmalsvektor gemäß

$$d(\lambda_i,\lambda_j) = \log p(X_i|\lambda_i) - \log p(X_i|\lambda_j) \quad (5)$$

bestimmt wird und der logarithmische Wahrscheinlichkeitsabstand zum zweiten Merkmalsvektor gemäß

$$d(\lambda_j,\lambda_i) = \log p(X_j|\lambda_j) - \log p(X_j|\lambda_i) \quad (6)$$

bestimmt wird, wobei zur Erzielung eines symmetrischen Abstandsmaßes der arithmetische Mittelwert zu

$$d(\lambda_j;\lambda_i) = \frac{1}{2}\left(d(\lambda_i,\lambda_j) + d(\lambda_j;\lambda_i)\right) \quad (7)$$

berechnet wird.

12. Verfahren nach Anspruch 11,
bei dem dieses hidden-Markov-Lautmodell (L,M,R) für die Modellierung des Lautes in mindestens beiden Sprachen nur verwendet wird, falls $d(\lambda_j;\lambda_i)$ eine festgelegte Schrankenbedingung erfüllt.

13. Verfahren nach einem der Ansprüche 1-7 zur Anpassung eines wie folgt gebildeten Mehrsprachen hidden-Markov-Lautmodelles:

a) ausgehend von mindestens einem ersten hidden-Markov-Lautmodell (L,M,R) für einen ersten Laut in einer ersten Sprache (SP,EN,DE) und von mindestens einem zweiten hidden-Markov-Lautmodell (L,M,R) für einen vergleichbar gesprochenen zweiten Laut in mindestens einer zweiten Spra-

che (DE,SP,EN), wird ein Poly-Phonem-Modell derart gebildet, daß die für die Modellierung des ersten und zweiten hidden-Markov-Lautmodelles (L,M,R) verwendeten Standardwahrscheinlichkeitsverteilungen (WD) bis zu einem festgelegten Abstandsschwellenwert, der angibt bis zu welchem maximalen Abstand zwischen zwei Standardwahrscheinlichkeitsverteilungen (WD) diese zusammengefügt werden sollen zu jeweils einer neuen Standardwahrscheinlichkeitsverteilung (WDP) zusammengefügt werden und lediglich die zusammengefügten Standardwahrscheinlichkeitsverteilungen das Poly Phonem Modell charakterisieren;

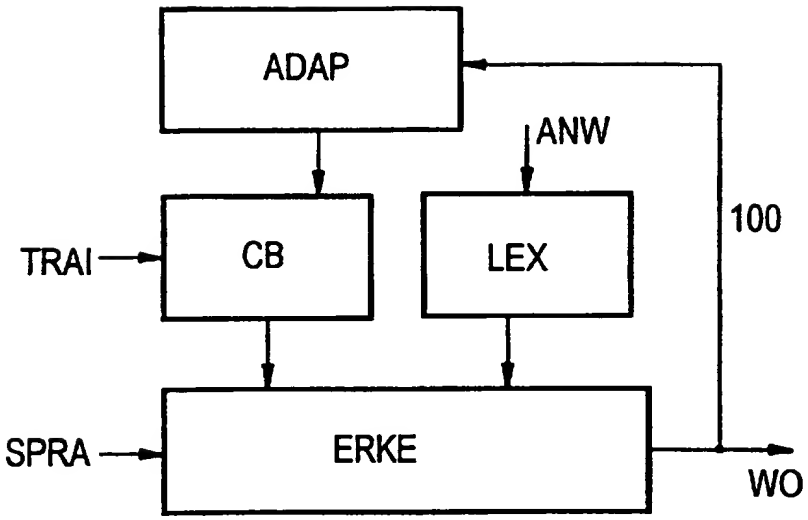
b) dieses Poly Phonem Modell wird für die Modellierung des Lautes in mindestens beiden Sprachen (DE,SP,EN) (L,M,R) verwendet.

14. Verfahren nach Anspruch 13,
bei dem als Abstandsschwellenwert 5 festgelegt wird.

15. Verfahren nach einem der vorangehenden Ansprüche,
bei dem hidden-Markov-Lautmodelle mit drei Zuständen verwendet werden, welche aus den Lautsegmenten Anlaut, Mittellaut und Ablaut gebildet werden.

16. Spracherkennungssystem, welches en Verfahren nach einem der vorangehenden Ansprüche ausführt

FIG 1



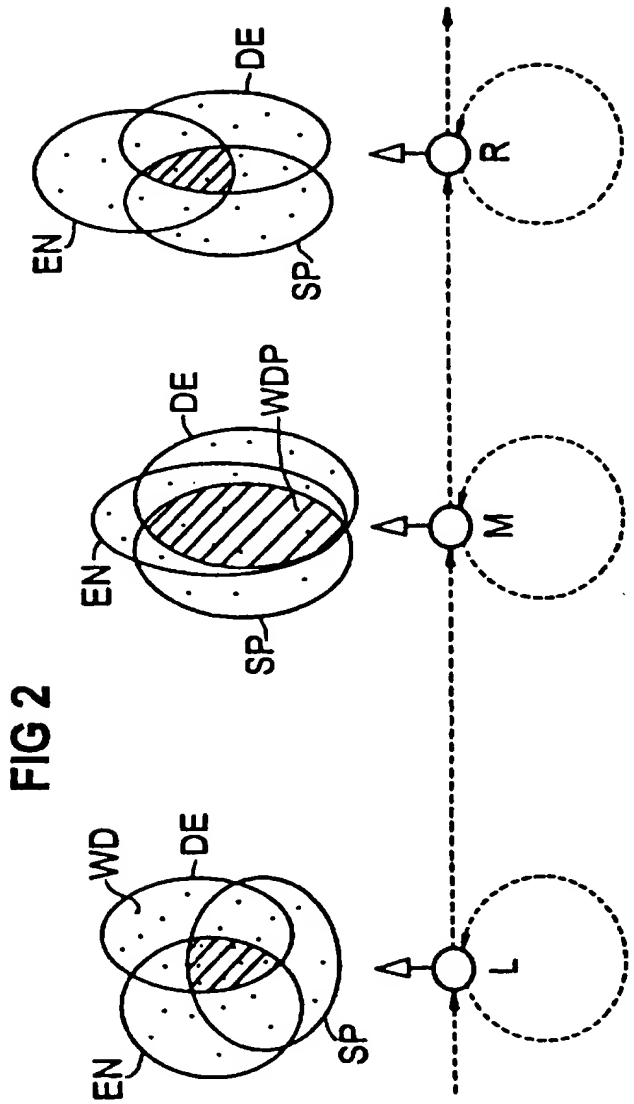


FIG 2

INTERNATIONAL SEARCH REPORT

Internatio pplication No
PCT/DE 97/02016

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 G10L5/06

According to International Patent Classification(IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	BUB U: "Task adaptation for dialogues via telephone lines" PROCEEDINGS ICSLP 96. FOURTH INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGE PROCESSING (CAT. NO.96TH8206), PROCEEDING OF FOURTH INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGE PROCESSING. ICSLP '96, PHILADELPHIA, PA, USA, 3-6 OCT. 1996, ISBN 0-7803-3555-4, 1996, NEW YORK, NY, USA, IEEE, USA, pages 825-828 vol.2, XP002053006 see paragraph 3.1 see paragraph 3.2 ---- -/--	1-7

[X] Further documents are listed in the continuation of box C.

[ ] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

3

Date of the actual completion of the international search	Date of mailing of the international search report
22 January 1998	04/02/1998

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Krembel, L

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Internatior pplication No
PCT/DE 97/02016

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
P,X	BUB U ET AL: "In-service adaptation of multilingual hidden-Markov-models" 1997 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING (CAT. NO.97CB36052), 1997 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, MUNICH, GERMANY, 21-24 APRIL 1997, ISBN 0-8186-7919-0, 1997, LOS ALAMITOS, CA, USA, IEEE COMPUT. SOC. PRESS, USA, pages 1451-1454 vol.2, XP002053007 see paragraph 2.1 see paragraph 2.3 ---	1-7
Y	NECIOGLU B F ET AL: "A BAYESIAN APPROACH TO SPEAKER ADAPTATION FOR THE STOCHASTIC SEGMENT MODEL" SPEECH PROCESSING 1, SAN FRANCISCO, MAR. 23 - 26, 1992, vol. 1, 23 March 1992, INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, pages 437-440, XP000341177 siehe Gleichung 15 ---	1-4
Y	PAUL D B ET AL: "THE LINCOLN LARGE-VOCABULARY STACK-DECODER HMM CSR" SPEECH PROCESSING, MINNEAPOLIS, APR. 27 - 30, 1993, vol. 2 OF 5, 27 April 1993, INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, pages II-660-663, XP000427876 siehe Seite 662 "Adaptation algorithm" ---	1-4
A	ALTO P ET AL: "ADAPTING A LARGE VOCABULARY SPEECH RECOGNITION SYSTEM TO DIFFERENT TASKS" SIGNAL PROCESSING THEORIES AND APPLICATIONS, BARCELONA, SEPT. 18 - 21, 1990, vol. 2, 18 September 1990, TORRES L;MASGRAU E; LAGUNAS M A, pages 1379-1382, XP000365815 see paragraph 3 see paragraph 5 ---	1
A	HSIAO-WUEN HON ET AL: "VOCABULARY LEARNING AND ENVIRONMENT NORMALIZATION IN VOCABULARY-INDEPENDENT SPEECH RECOGNITION" SPEECH PROCESSING 1, SAN FRANCISCO, MAR. 23 - 26, 1992, vol. 1, 23 March 1992, INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, pages 485-488, XP000341189 siehe Absatz "Vocabulary-Bias Training" --- -/--	1

3

INTERNATIONAL SEARCH REPORT

Internatior pplication No
PCT/DE 97/02016

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	HOFSTETTER E M ET AL: "TECHNIQUES FOR TASK INDEPENDENT WORD SPOTTING IN CONTINUOUS SPEECH MESSAGES*" SPEECH PROCESSING 2, AUDIO, NEURAL NETWORKS, UNDERWATER ACOUSTICS, SAN FRANCISCO, MAR. 23 - 26, 1992, vol. 2, 23 March 1992, INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, pages 101-104, XP000356947 siehe Absatz "4. Speaker Adaptation / Unsupervised reestimation of a linear filtering"	1

3

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONALER RECHERCHENBERICHT

Internatio . Aktenzeichen
PCT/DE 97/02016

A. KLASSIFIZIERUNG DES ANMELDUNGSGEGENSTANDES
IPK 6 G10L5/06

Nach der Internationalen Patentklassifikation (IPK) oder nach der nationalen Klassifikation und der IPK

B. RECHERCHIERTE GEBIETE

Recherchierter Mindestprüfstoff (Klassifikationssystem und Klassifikationssymbole)
IPK 6 G10L

Recherchierte aber nicht zum Mindestprüfstoff gehörende Veröffentlichungen, soweit diese unter die recherchierten Gebiete fallen

Während der internationalen Recherche konsultierte elektronische Datenbank (Name der Datenbank und evtl. verwendete Suchbegriffe)

C. ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
P,X	BUB U: "Task adaptation for dialogues via telephone lines" PROCEEDINGS ICSLP 96. FOURTH INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGE PROCESSING (CAT. NO.96TH8206), PROCEEDING OF FOURTH INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGE PROCESSING. ICSLP '96, PHILADELPHIA, PA, USA, 3-6 OCT. 1996, ISBN 0-7803-3555-4, 1996, NEW YORK, NY, USA, IEEE, USA, Seiten 825-828 vol.2, XP002053006 siehe Absatz 3.1 siehe Absatz 3.2 --- -/--	1-7

[X] Weitere Veröffentlichungen sind der Fortsetzung von Feld C zu entnehmen

[ ] Siehe Anhang Patentfamilie

* Besondere Kategorien von angegebenen Veröffentlichungen :

"A" Veröffentlichung, die den allgemeinen Stand der Technik definiert, aber nicht als besonders bedeutsam anzusehen ist

"E" älteres Dokument, das jedoch erst am oder nach dem internationalen Anmeldedatum veröffentlicht worden ist

"L" Veröffentlichung, die geeignet ist, einen Prioritätsanspruch zweifelhaft erscheinen zu lassen, oder durch die das Veröffentlichungsdatum einer anderen im Recherchenbericht genannten Veröffentlichung belegt werden soll oder die aus einem anderen besonderen Grund angegeben ist (wie ausgeführt)

"O" Veröffentlichung, die sich auf eine mündliche Offenbarung, eine Benutzung, eine Ausstellung oder andere Maßnahmen bezieht

"P" Veröffentlichung, die vor dem internationalen Anmeldedatum, aber nach dem beanspruchten Prioritätsdatum veröffentlicht worden ist

"T" Spätere Veröffentlichung, die nach dem internationalen Anmeldedatum oder dem Prioritätsdatum veröffentlicht worden ist und mit der Anmeldung nicht kollidiert, sondern nur zum Verständnis des der Erfindung zugrundeliegenden Prinzips oder der ihr zugrundeliegenden Theorie angegeben ist

"X" Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann allein aufgrund dieser Veröffentlichung nicht als neu oder auf erfinderischer Tätigkeit beruhend betrachtet werden

"Y" Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann nicht als auf erfinderischer Tätigkeit beruhend betrachtet werden, wenn die Veröffentlichung mit einer oder mehreren anderen Veröffentlichungen dieser Kategorie in Verbindung gebracht wird und diese Verbindung für einen Fachmann naheliegend ist

"&" Veröffentlichung, die Mitglied derselben Patentfamilie ist

Datum des Abschlusses der internationalen Recherche	Absendedatum des internationalen Recherchenberichts
22.Januar 1998	04/02/1998

Name und Postanschrift der Internationalen Recherchenbehörde	Bevollmächtigter Bediensteter
Europäisches Patentamt, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Krembel, L

Formblatt PCT/ISA/210 (Blatt 2) (Juli 1992)

INTERNATIONALER RECHERCHENBERICHT

Internatic s Aktenzeichen
PCT/DE 97/02016

C.(Fortsetzung) ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
P,X	BUB U ET AL: "In-service adaptation of multilingual hidden-Markov-models" 1997 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING (CAT. NO.97CB36052), 1997 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, MUNICH, GERMANY, 21-24 APRIL 1997, ISBN 0-8186-7919-0, 1997, LOS ALAMITOS, CA, USA, IEEE COMPUT. SOC. PRESS, USA, Seiten 1451-1454 vol.2, XP002053007 siehe Absatz 2.1 siehe Absatz 2.3 ---	1-7
Y	NECIOGLU B F ET AL: "A BAYESIAN APPROACH TO SPEAKER ADAPTATION FOR THE STOCHASTIC SEGMENT MODEL" SPEECH PROCESSING 1, SAN FRANCISCO, MAR. 23 - 26, 1992, Bd. 1, 23.März 1992, INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, Seiten 437-440, XP000341177 siehe Gleichung 15 ---	1-4
Y	PAUL D B ET AL: "THE LINCOLN LARGE-VOCABULARY STACK-DECODER HMM CSR" SPEECH PROCESSING, MINNEAPOLIS, APR. 27 - 30, 1993, Bd. 2 OF 5, 27.April 1993, INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, Seiten II-660-663, XP000427876 siehe Seite 662 "Adaptation algorithm" ---	1-4
A	ALTO P ET AL: "ADAPTING A LARGE VOCABULARY SPEECH RECOGNITION SYSTEM TO DIFFERENT TASKS" SIGNAL PROCESSING THEORIES AND APPLICATIONS, BARCELONA, SEPT. 18 - 21, 1990, Bd. 2, 18.September 1990, TORRES L;MASGRAU E; LAGUNAS M A, Seiten 1379-1382, XP000365815 siehe Absatz 3 siehe Absatz 5 ---	1
A	HSIAO-WUEN HON ET AL: "VOCABULARY LEARNING AND ENVIRONMENT NORMALIZATION IN VOCABULARY-INDEPENDENT SPEECH RECOGNITION" SPEECH PROCESSING 1, SAN FRANCISCO, MAR. 23 - 26, 1992, Bd. 1, 23.März 1992, INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, Seiten 485-488, XP000341189 siehe Absatz "Vocabulary-Bias Training" --- -/--	1

3

INTERNATIONALER RECHERCHENBERICHT

Internatio ı Aktenzeichen
PCT/DE 97/02016

C.(Fortsetzung) ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
A	HOFSTETTER E M ET AL: "TECHNIQUES FOR TASK INDEPENDENT WORD SPOTTING IN CONTINUOUS SPEECH MESSAGES*" SPEECH PROCESSING 2, AUDIO, NEURAL NETWORKS, UNDERWATER ACOUSTICS, SAN FRANCISCO, MAR. 23 - 26, 1992, Bd. 2, 23.März 1992, INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, Seiten 101-104, XP000356947 siehe Absatz "4. Speaker Adaptation / Unsupervised reestimation of a linear filtering"	1

3